103864
1

Multimodal Gestural Control Using On-body Sensors – Final Report

Thad Starner
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0760 USA
+1 (404) 385-0816
{thad}@cc.gatech.edu

February 14, 2007

1 Introduction

Mobile phones are now the dominant computing device on the planet. In the developing world, mobile phones are beginning to take over the traditional role of the PC due to the resource limitations in remote areas. In technologically developed markets, phone handsets offer desktop-like applications for convenience. However, mobile interfaces can not assume the traditional mouse, keyboard, and screen of the desktop computer. New interfaces are necessary for users that are "on-the-go," and there is an opportunity to re-define what the "standard" computer interface means.

Electronic devices are becoming smaller; however, the same trend can not continue for the devices' user interfaces. For instance, due to the physical size of the human hand, buttons cannot be made any smaller and still be reliably pressed or seen (see Figure 1). Age and illness exacerbate this effect.

Traditional button interfaces may also be inappropriate in situations where visual attention is limited or where user contact may be unhygienic or messy. For example, manipulating a car radio or mobile phone often requires visual attention to search for the buttons, which may be dangerous while driving. A surgeon may not want to touch medical equipment while operating but still wish to control devices such as microscopes or medical monitors. Similarly, car mechanics may not wish to touch their computerized diagnostic equipment while their hands are covered in grease. Non-contact gesture recognition can overcome many of these limitations but has had limited application in mobile computing to date. [Interestingly, many laptops and mobile music players now have built-in accelerometers to detect free-fall. When a "dropped the laptop" gesture is detected, the laptop parks its hard drive! With programming, these accelerometers can be re-purposed for some of the ideas presented here.]

Gestural input provides an alternative input method for computing that can help address these issues. Gesture interfaces can be integrated into the smallest of devices, and many gestures can be performed while physically moving. In addition, gesture interfaces can enable users with physical disabilities to use computing devices that were formerly inaccessible.

Since the beginning of this project, the Nintendo Wii has demonstrated that gesture interfaces are appropriate (and popular) for consumer electronics. Unfortunately, the Wii's gesture technology is not appropriate for the mobile domain. Apple's new iPhone begins to address mobile gesture interfaces. The multi-touch gesture recognition system that Apple is pursuing leverages past HCI research and is an innovation for the electronics industry. However, it has limited applicability. It assumes that the user's full attention is on the interface, the user is holding the display, and that the user has fine motor control. We believe these restrictions are unrealistic for more on-the-go tasks (such as controlling audio playing and recording, triggering image capture, entering notes in a meeting or at a chance encounter, on-the-go web browsing, and controlling appliances in the environment).

This project continues to examine mobile gesture interfaces and infrastructure that may be used with the next generation PC. The Gesture Pendant 2 and Gesture Watch are devices, worn as a broach or wristwatch respectively, that allow the user to make gestures to control appliances near him. Bluesense creates a Bluetooth sensor toolkit that

Figure 1: MP3 players and phone headsets could shrink to the scale of a hearing aid but would require an alternative to button interfaces due to their size.

allows rapid development of new physical interface devices. The Georgia Tech Gesture Toolkit allows HCI (human computer interaction) researchers to develop gesture-based interfaces with minimal knowledge of pattern recognition. We demonstrate the use of this infrastructure in "Mobile Dance Revolution," a portable, mobile phone version of the popular "Dance Dance Revolution" arcade game. A final, but key, result is our work in text entry, which examines replacements for the desktop keyboard while the user in mobile.

1.1 Interaction with ETRI

Jung Soo Kim, one of our graduate students, worked at ETRI last summer to help transfer the Georgia Tech Gesture Toolkit (GT2K). He built on the work started by Seung Yon Lee's (a PhD student) visit in Spring 2006. (Seung Yon had brought early prototypes of Georgia Tech's hardware to ETRI.) Jung Soo expects to return to ETRI Summer 2007.

ETRI researchers Dong Woo Lee and Jin Ho Yoo visited Georgia Tech for six months to work on ETRI's PMG and integrate our prototypes with that effort.

Seung Yon and Jung Soo visited ETRI February 1st to demonstrate the Gesture Pendant 2, Gesture Watch, and Mobile Dance Revolution prototypes as part of the ETRI demonstration for the Korean government. We provided slides, images, and movies for the government presentation. Last, we delivered the Bluesense documentation and reports from our mobile text entry studies.

2 Gesture Pendant 2

In a previous project, we created the Gesture Pendant, a mobile gesture interface designed to allow the wearer to control appliances in the house via hand gestures. For example, the wearer can simply raise or lower a flattened hand to control the dimming of a light. Similarly, the user could control the volume of the stereo by raising or lowering a pointed finger. By putting the sensing and computing on the body, this same pendant can be used to control things in the office, in the car, on the sidewalk, or at a friend's house. The Gesture Pendant consisted of a small camera worn as a part of a necklace or pin. The camera is ringed by IR LEDs and has a IR pass filter over the lens. The result is a camera that is sensitive to the user's hands, even in the dark.

While the Gesture Pendant was demonstrated successfully to hundreds of users and was invited to exhibitions and demonstrations, it is not practical. Due to the system's reliance on infrared light, the system can confuse sunlight and bright indoor halogen lights for the user's hand.

After some experimentation (the Freedigiter work from last year), we determined that infrared proximity sensors might be used instead of the infrared vision system to detect the user's hand. These sensors automatically adjust for ambient IR levels. Thus, we re-examined the Gesture Pendant prototype and embarked on making a new version using these sensors.

At the end of July, we provided ETRI with a prototype of the Gesture Pendant 2. This prototype uses four of the proximity sensors to create a new type of mobile gesture recognition system. Each proximity sensor can sense extended fingers passing over it, and we believe the set of four will allow more general gestures, similar in spirit to the original Gesture Pendant. For example, the four proximity sensors could be mounted in a vertical line. The user can extend a finger and slide it up and down this line of sensors to indicate the raising or lowering of the volume of a radio. Two fingers raised or lowered might indicate the brightening or dimming of the lights in a house. Three fingers might be used to control the switching of channels on a television.

Alternatively, the proximity sensors might be arranged in a circular pattern. A hand passing over the circle in a clockwise pattern might indicate that a DVD player should fast forward. Similarly, a counterclockwise motion might indicate that the DVD player should rewind. Using two fingers would indicate that these actions should occur at 2X normal speed. Three fingers would indicate 3X, and so forth.

Note that this Gesture Pendant 2 system could be integrated into small mobile consumer devices. For example, the jog wheel of the iPod might be replaced by such a system. Similarly, the Gesture Pendant 2 system could be embedded in home appliances. For example, a VCR or DVD player could use the proximity sensors so as to provide a non-contact interface.

In order to enable ETRI with the maximal ability to experiment with these ideas, we provided a prototype system that could be reconfigured as desired. A complete system (a PC, a USB-based analog to digital converter, and four proximity sensors) was provided. Each proximity sensor could be mounted on a small velcro mount to create the topology desired. In addition, we provided a scripting gesture recognition system that allows the creation and composition of new gestures. These gestures can then be associated with actions to perform on a PC (such as controlling a media player). An example demonstration was included using Linux, and all software was included on a bootable CD-ROM.

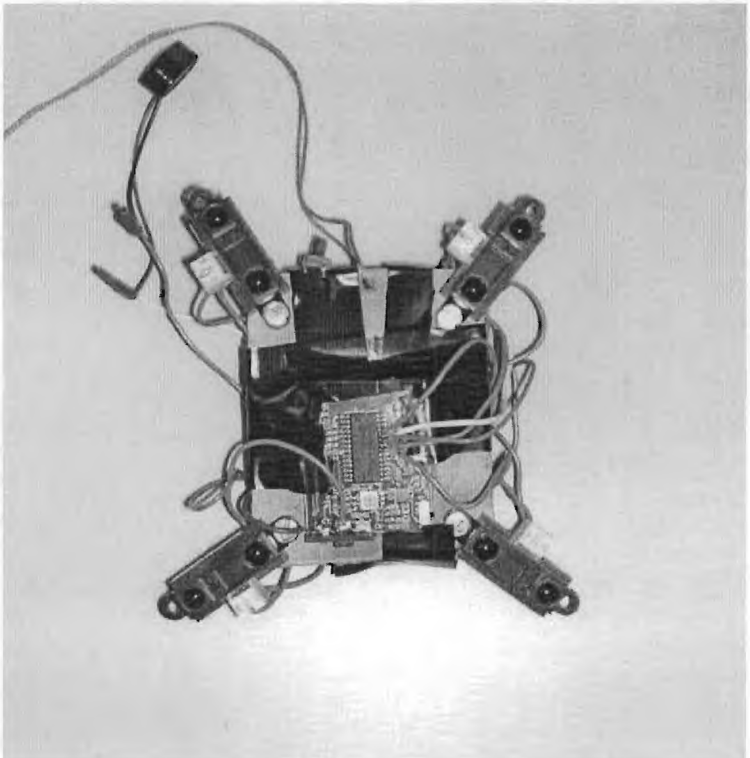

Figure 2: A rough prototype of the Gesture Pendant 2 which uses Bluetooth to communicate to the environment.

We have continued working on the prototype and now have a version where the infrared sensors transmit their state over Bluetooth to a remote computer (see Figure 2 - 6). We have designed a simple visualization interface to show the state of the Gesture Pendant 2 as it is being trained with GT2K for recognizing new gestures (Figure 5).

Recently, we have re-packaged the prototype into a form more appropriate for user testing. The prototype senses the user's gestures, and gesture recognition occurs on a desktop PC using GT2K. Once the proper command is de-

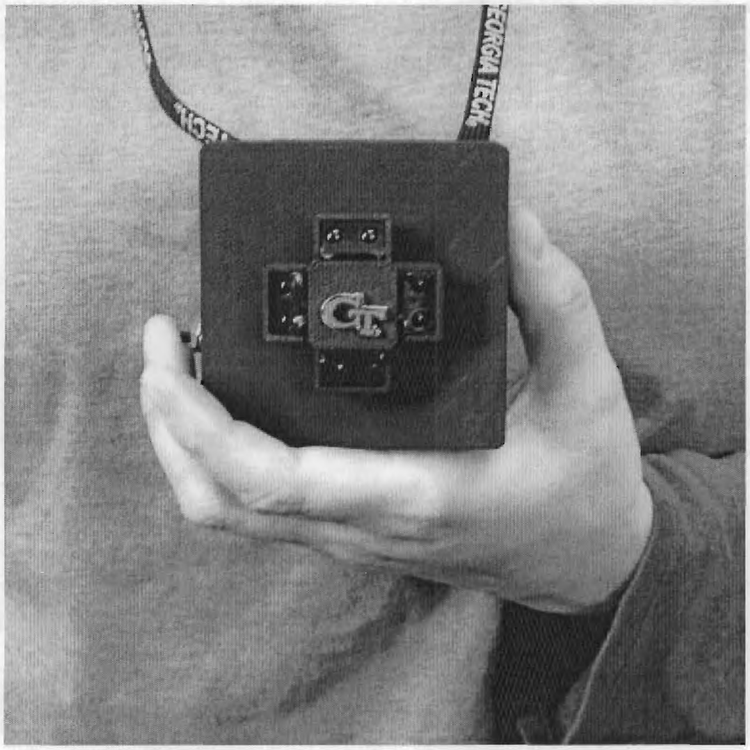

Figure 3: A second revision of the Gesture Pendant 2 (prototype case produced on Dimension 3D printer).

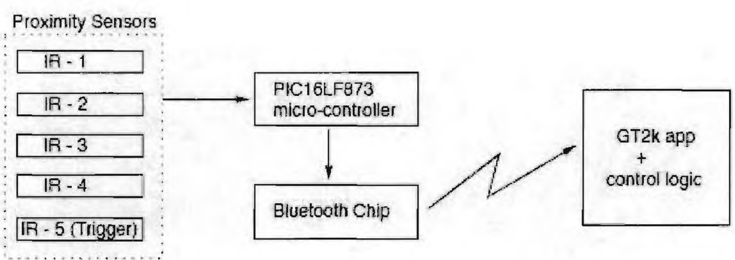

Figure 4: Gesture Pendant 2 system diagram.

termined, the appropriate instruction (for example, play or fast forward) is sent to the target device (a DVD player or television) through an infrared control module connected to the PC (Figure 6). The IR control modules provide universal remote control functionality for all of our projects that control consumer appliances.

Encouraged by results with the Gesture Pendant 2, we decided to determine if the hardware could be made small enough to fit into a wristwatch form factor. Our original rough prototype can be seen in Figure 7. Using this prototype, we have determined that the user can perform relatively sophisticated control gestures. Figure 8 shows a refined prototype we are preparing for more formal user testing. However, even with the more refined version, the proximity sensors require too much space for embedding in an off-the-shelf wristwatch. We have found a new proximity sensor kit on the market that requires approximately a tenth of the volume of the current sensors, and we are experimenting with it.

We have connected the Gesture Pendant 2 and Gesture Watch to Jin Ho's prototype of ETRI's Personal Media Gateway (PMG). Unfortunately, we had to use version 1 of GT2K to get the speed we desire (as opposed to the more

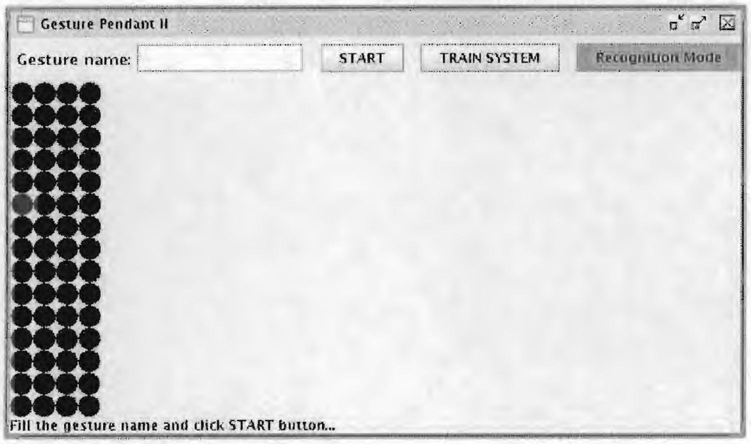

Figure 5: A visualization displays the state of the Gesture Pendant 2's IR sensors as the system is being trained to recognize new gestures with GT2K.

recent user-friendly Java-based version). We are investigated faster algorithms to include in the next version of GT2K.

3 Bluesense: Bluetooth accelerometers and sensing platform

Much of our previous work leveraged our Bluetooth accelerometers, and we have received many outside requests from other researchers wishing to buy them. These systems, though one of the first devices of their kind when designed, are now getting out of date. The Bluesense platform re-create the Bluetooth accelerometers in a smaller form factor while requiring less power (see Figure 9). These accelerometers are more sensitive and have a higher dynamic range. However, the main goal of the Bluesense platform is to "lower the barrier" for interface researchers to create wireless interfaces. With a basic understanding of electronics, an interface researcher is able to attach new sensors to the Bluesense board and receive readings remotely over Bluetooth.

Bluesense has been used on several projects, including the Gesture Pendant 2 and Gesture Watch above. However, the system remains too complicated. One major difficulty is that the microcontroller in the Cambridge Silicon Radio (CSR) board must dedicate itself mostly to controlling the Bluetooth radio. When we multiplex this microcontroller to read the interface to our sensors, the system can only transmit readings at 3 Hz. Fortunately, the CSR board also includes a DSP co-processor. We have re-purposed the DSP to read the sensor interface and have achieved acceptable frame rates. Unfortunately, the DSP must be programmed in assembly, which is a significant barrier for many interface designers. We continuing to update a application programming interface to shield the designer from these details. In addition, we have created documentation on how to use the Bluesense platform to create Bluetooth keyboards and mice using the HID profile. The HID profile allows the installed base of Bluetooth compatible computers, PDAs, and mobile phones to use our prototype texting and pointing interfaces without the need for software drivers. A class project recently redesigned the Twiddler one-handed chording keyboard so that it could connect to computers (and mobile phones) using Bluesense and its HID package.

We have created an entertaining demonstration of the Bluesense platform: Mobile Dance Revolution (MDR). MDR is a version of the game "Dance Dance Revolution" for a mobile phone. Bluesense accelerometers are laced into the player's shoes. The user performs "dance" steps synchronized to the game's interface in order to score points. Given recent demonstrations, many people find the game quite compelling, and it suggests that mobile gesture games might be appropriate for mobile phone users (for example, extending some of the game play ideas of the popular Nintendo DS).

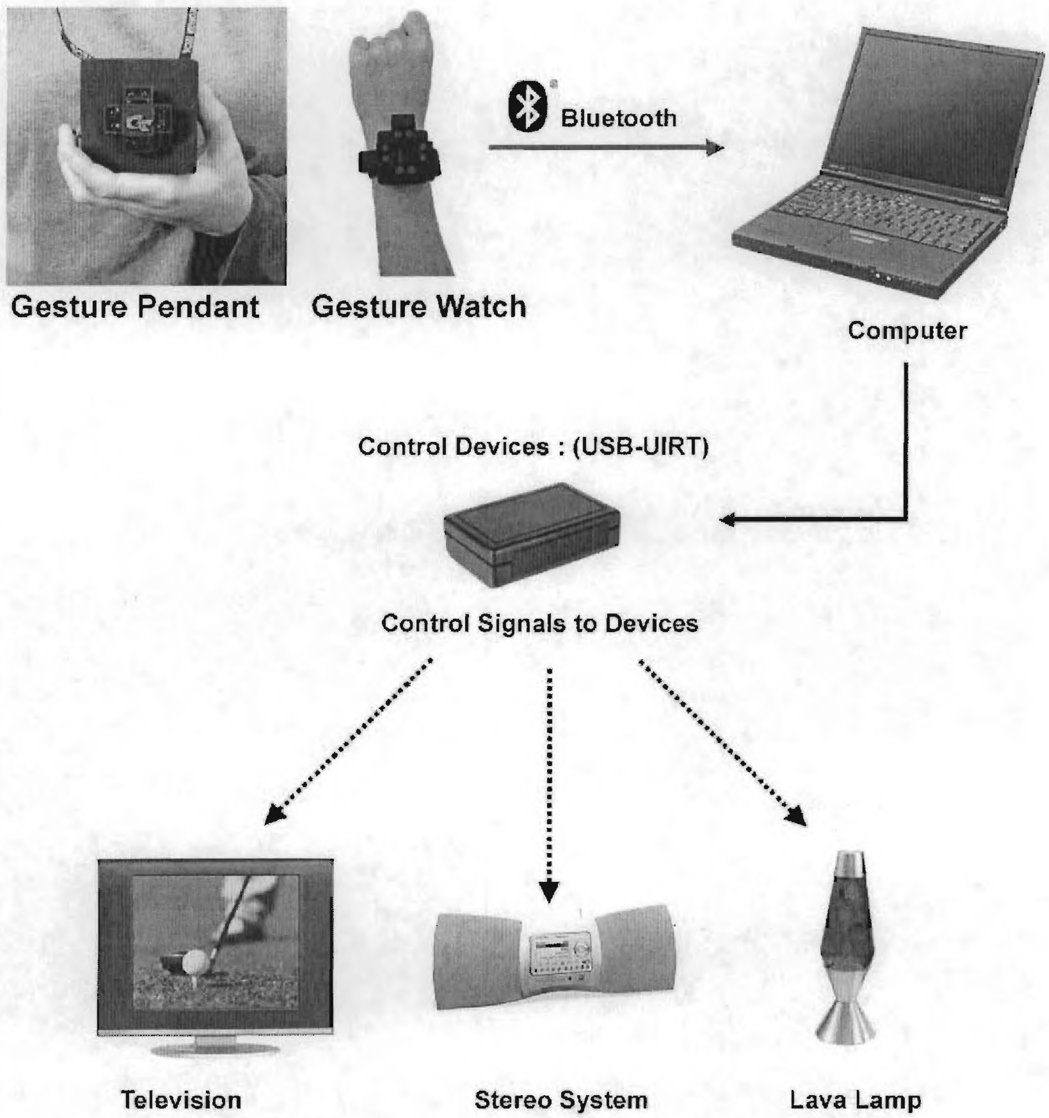

Figure 6: In simulating the final Gesture Pendant 2 and Gesture Watch interfaces for user testing, an IR command module (connected to a PC through USB) issues control commands for appliances. In practice, the computation and infrared control would be embedded in the Gesture Pendant 2 or Gesture Watch themselves.

4 Georgia Tech Gesture Toolkit (GT2K)

One of our goals is to develop the Georgia Tech Gesture Toolkit (GT2K) so that interface designers without a background in HMMs or pattern recognition can use the system to create gestural interfaces. We also wish to continue to develop more advanced tools for GT2K for the gesture recognition specialist.

As mentioned previously, Jung Soo successfully transferred an earlier version of GT2K to ETRI during his summer internship. Version 2 of the system is now functional and is being used in our current projects with ETRI. Jung Soo has been recording the problems ETRI and our students have encountered in using GT2K and has been steadily improving

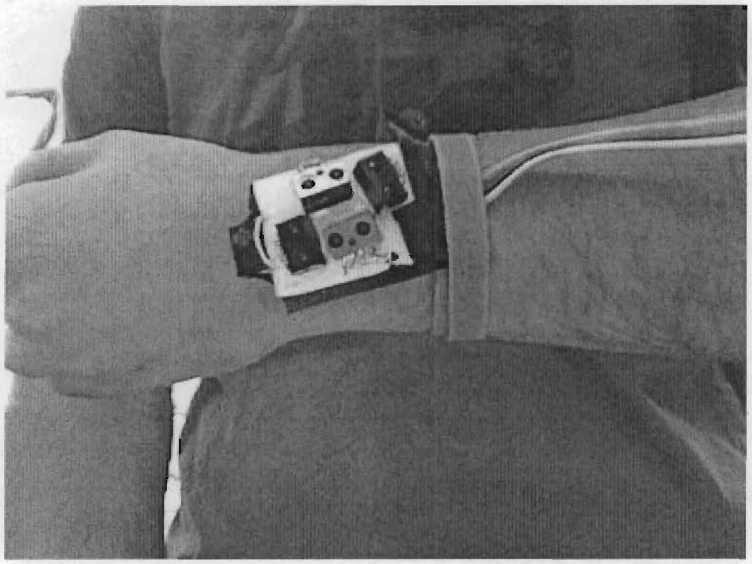

Figure 7: Rough prototype of the Gesture Watch.

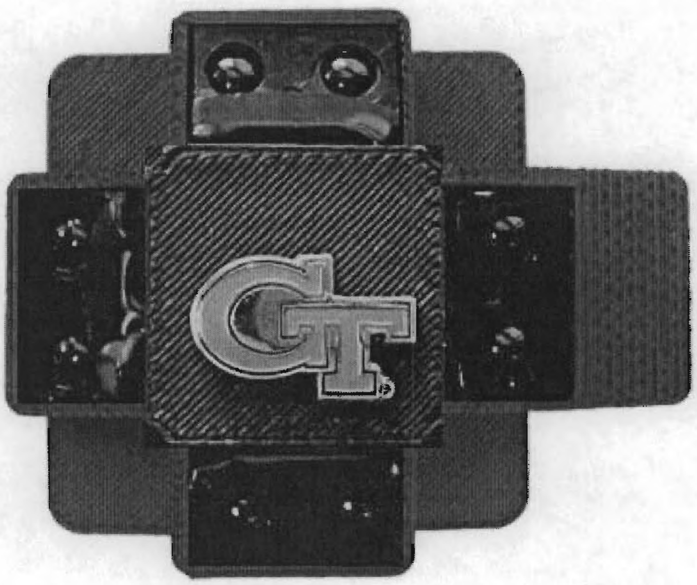

Figure 8: Gesture Watch ready for user testing.

the system. Given the substantive improvements, we have renamed the system the "Gesture and Activity Recognition Toolkit (GART)" and are in the process of publishing a paper at HCII describing the system (an abstract is included in the supplemental information). Figure 10 shows a system diagram of GART. We have adopted Bugzilla to track bugs and improvements in the system as we begin to release GART publicly. Currently, GART is being used as part of our project classes at Georgia Tech.

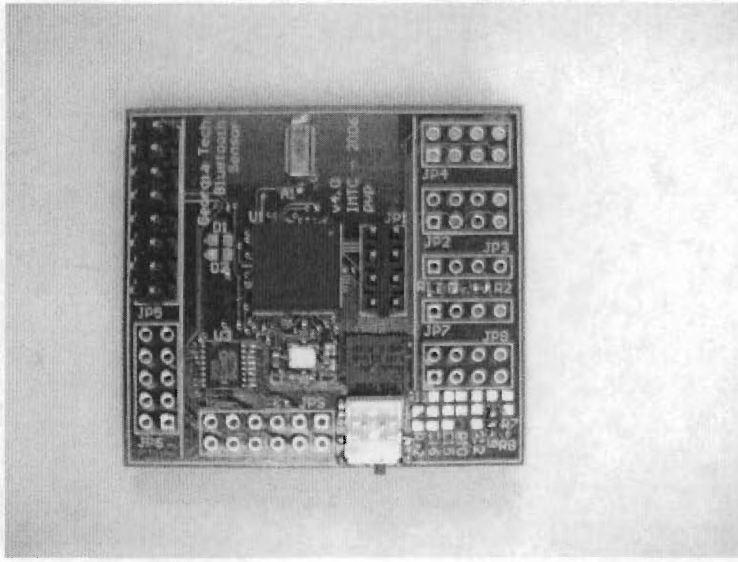

Figure 9: Prototype BlueSense sensor and radio board.

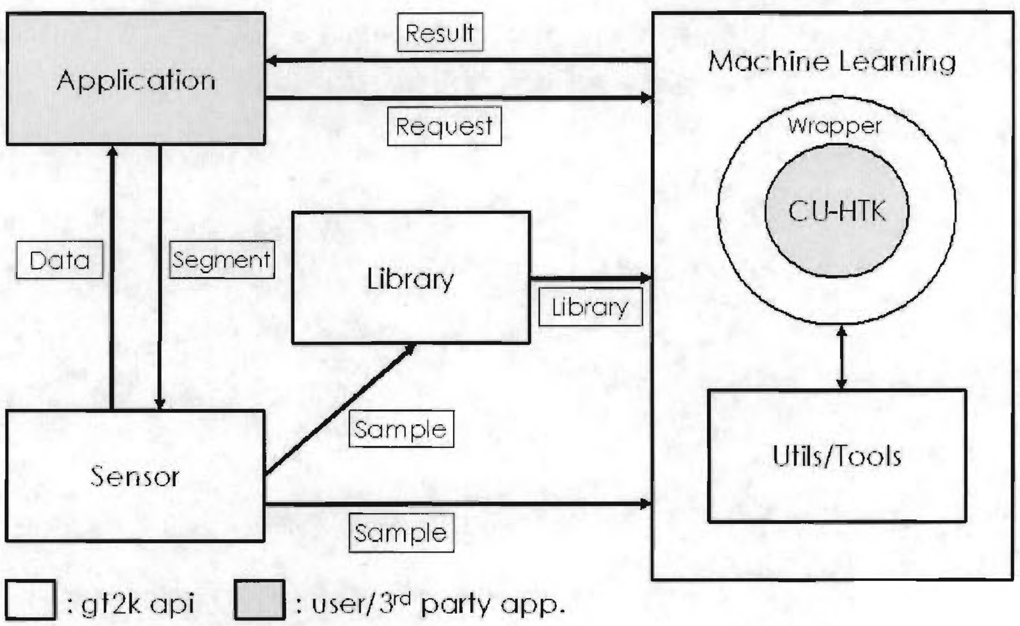

Figure 10: System diagram of the Gesture and Activity Recognition Toolkit (GART).

5 Mobile Text Entry

Our recent work on mobile text entry has led to the surprising conclusion that users can reach desktop-level typing speeds on mobile-phone sized keyboards. Expert typists using both the Twiddler and mini-QWERTY (RIM Blackberry) style keyboards can exceed 60 words per minute. Our initial work on a Korean version of the Twiddler suggests that equivalent speeds (or greater) might be reached for Hangul.

Our initial studies had subjects use the Twiddler and mini-QWERTY keyboards while they were sitting at a desk. We noticed that mini-QWERTY users visually focused on the keyboard while using it. While mobile, a user can not afford such visual attention. We suspected that the mini-QWERTY keyboard might not be a good keyboard while the user is moving. To test this hypothesis, we had expert mini-QWERTY typists attempt to type "blind" without seeing the keyboard (keeping it placed underneath the desk). Their typing rates went from 60 wpm to 45 wpm and their error rates increased from 6% per character to 15% per character. Meanwhile, Twiddler experts were unaffected by typing blind.

By comparing Twiddler novice learning rates to desktop QWERTY learning rates, we suspect (but have not proven) that a user who had never used a Twiddler or a QWERTY keyboard before would learn the Twiddler 2-3 times faster than the mini-QWERTY keyboard. In addition, the Twiddler requires less moving parts and can be designed into a smaller form factor than the mini-QWERTY. Thus, if a mobile keyboard was introduced to a population unfamiliar with the western Latin alphabet keyboard (for example, in the developing world) or if the language was not suitable for touch typing on a QWERTY keyboard, the Twiddler could be a very good alternative.

Papers detailing the results of our studies are included in the supplemental material.

6 Conclusion

We have described our research on mobile gesture interfaces and provided prototypes of our GT2K software, Gesture Pendant 2, Gesture Watch, and Bluesense hardware to ETRI's laboratories. We have completed our mobile text entry survey and testing and have included documentation on the results in our supplemental material. We continue to work with ETRI researchers in improving our demonstrations and integrating our work with ETRI's prototypes.